A review of *Archiving for the future: Simple steps for archiving language documentation collections*

Victoria Baldwin

Archiving for the future: Simple steps for archiving language documentation collections (Kung et al. 2020) (henceforth AFTF) is an online course that sets out to comprehensively prepare those engaged in language documentation projects to meet a framework of archiving for language documentation which has been developed over the last two decades. To effectively evaluate it, it is necessary to briefly explore key aspects of that framework.

In response to changing technology, Bird and Simons (2003) set out a blueprint to ensure the coherence, quality, and longevity of digitised data in language documentation. As the volume of documented material increased, Nathan & Austin (2004) advocated creating detailed metadata to enhance accessibility to an increasing "quicksand" of information. Holton (2012) showed that language archives have scientific and pedagogical applications beyond linguistics and therefore should be designed with these broader uses in mind. Nathan (2014) emphasised the need for access by communities to materials about their language and culture. This idea has since been established in doctrines such as the FAIR Principles (Wilkinson et al. 2016), and the CARE Principles for Indigenous Data Governance (Carroll et al. 2020).

On this basis, substantial infrastructure has emerged with archives such as the Endangered Languages Archive (ELAR) and the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). Specialist software, like ELAN and Lameta, has also been developed. This documentary infrastructure aims to ensure that data for archiving is good quality, technologically future-proofed and well supported by metadata. It must be accessible to a variety of users and adhere to established ethical principles.

AFTF attempts to integrate these approaches and infrastructure into a single, accessible training resource. The training is divided into three phases, plus an introduction and conclusion. The introduction establishes its purpose and goals and presents the structure for subsequent sections. The introduction goes on to provide an overview of language documentation since the time of linguistic pioneers like Franz Boas, much in the same vein as the manner described by Henke and Berez-Kroeker (2016). A clear visual (see Figure 1) allows participants to develop a picture of how the course will proceed. Participants are then invited to carry out a kind of "pre-task" activity, aimed at eliciting prior knowledge of collecting and organising data by considering their own collections and asking them to create their own inventory. The aim of such a task is to contextualise the topics that will be covered in the course and prepare participants to learn. The introduction concludes with a summary of key vocabulary and provides suggestions for further reading. This basic structure of introducing concepts, reflection, vocabulary, and reading is repeated throughout subsequent sections.

Simple steps for archiving language documentation collections

PHASE 1 BEFORE	STEP 1 choose your archive	STEP 2 name your files	STEP 3 pick enduring file formats	STEP 4 understand metadata
PHASE 2 DURING	STEP 5 collect metadata	STEP 6 evaluate your files		
PHASE 3 AFTER	STEP 7 prepare your deposit	STEP 8 practice progressive archiving	STEP 9 describe your collection	

Figure 1 Phases One to Three, from Introduction to Archiving for the Future: Simple Steps for Archiving Language Documentation Collections

Phase One covers the planning stage of language documentation projects, guiding language documenters through the contemporary framework of the discipline at every step of their project. Step One introduces some well-known archives, including many belonging to The Digital Endangered Languages and Musics Archives Network (DELAMAN) and others. Step Two discusses the organisation of files, while Step Three delves into the kind of technical and media quality issues raised by Jukes (2011), for example, opting for open formats rather than proprietary ones, or planning video for streaming rather than archiving large video files that are cumbersome to download. In this way, accessibility is built into the archiving process. Step Four introduces the concept of metadata, using visuals to simplify concepts. (See Figure 2.)
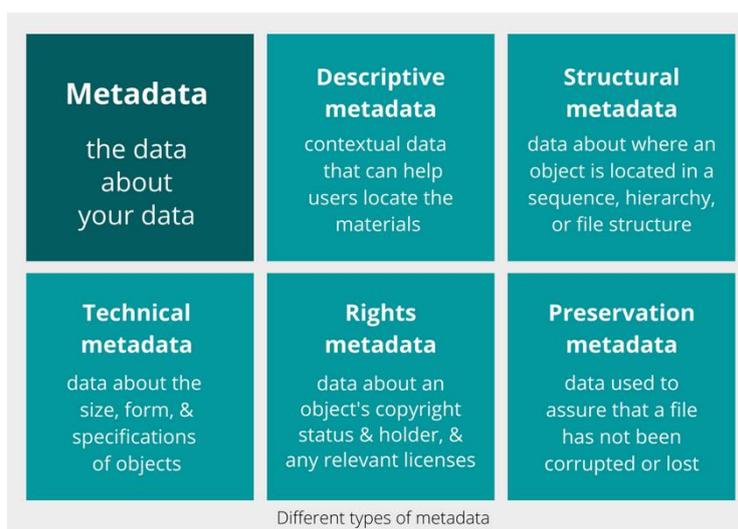

| **Metadata** the data about your data | **Descriptive metadata** contextual data that can help users locate the materials | **Structural metadata** data about where an object is located in a sequence, hierarchy, or file structure |
| **Technical metadata** data about the size, form, & specifications of objects | **Rights metadata** data about an object's copyright status & holder, & any relevant licenses | **Preservation metadata** data used to assure that a file has not been corrupted or lost |

Different types of metadata

Figure 2 Visual aid for understanding metadata, from Section Two, Phase One, Step Four

Phase Two develops the discussion around metadata, with Step Five introducing various tools and approaches used to create and manage metadata in the field. For example, advocating the use of software such as Lameta or SayMore for collating and tracking

metadata, or inviting documenters to consider issues around access such as copyright or licensing. Step Six discusses appraisal of collected materials, including sensitivity to Traditional Knowledge, and other forms of intangible cultural heritage. For example, communities having strong views toward certain cultural artefacts which make their deposit in an archive problematic. AFTF makes useful and ethical suggestions on how to resolve such issues. Examples of issues discussed are shown in Figure 3.
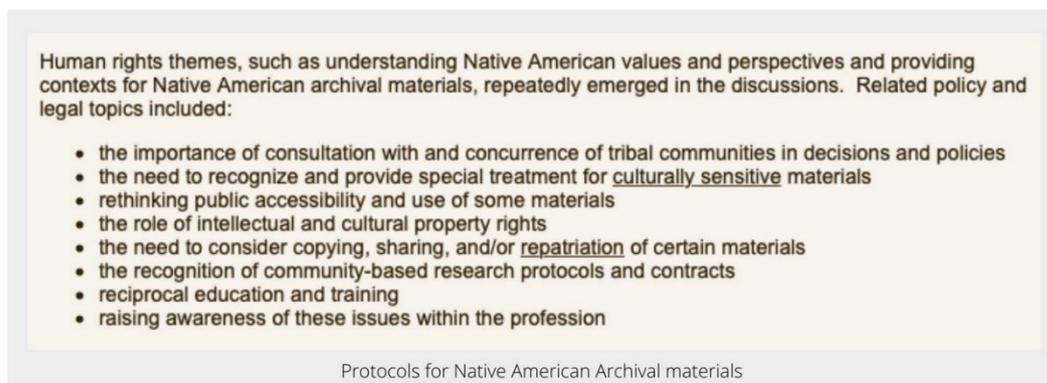

Human rights themes, such as understanding Native American values and perspectives and providing contexts for Native American archival materials, repeatedly emerged in the discussions. Related policy and legal topics included:

- the importance of consultation with and concurrence of tribal communities in decisions and policies
- the need to recognize and provide special treatment for culturally sensitive materials
- rethinking public accessibility and use of some materials
- the role of intellectual and cultural property rights
- the need to consider copying, sharing, and/or repatriation of certain materials
- the recognition of community-based research protocols and contracts
- reciprocal education and training
- raising awareness of these issues within the profession

Protocols for Native American Archival materials

Figure 3 Considerations in Archiving Traditional Knowledge, from Step Six

Phase Three discusses deposit into an archive. Step Seven explores strategies for arranging materials into coherent bundles of data. Suggestions include opting for a flat file structure, rather than a nested one since it is easier for archive users to navigate. It might also encourage depositors to arrange files by association (e.g., all files relating to a particular recorded event) rather than by type (i.e., all video files in one folder, all transcripts in another and leaving archive users to muddle their way through). Not only does this keep materials coherent, but it also increases accessibility. Step Eight introduces the concept of progressive archiving, as discussed by Nathan (2013), where essentially materials are added to a collection one piece at a time, rather than as one unwieldy mass of data that must be organised before the entire collection can be archived. This is helpful both as a means to maintain quality but also as a tool for managing workflows around large volumes of data. Step Nine discusses issues of accessibility through the production of a collection guide, as expanded most recently by Sullivant (2020). This essentially is a document detailing the key features and content of the archived collection with the intent of making it easier for people to access and make use of that content.

The conclusion shares citations, acknowledgements, and attributions. Participants may also request a certificate of completion.

AFTF is clearly built around the core framework of contemporary language documentation. As an educational tool, it provides a thorough introduction to the processes and tools of language documentation. The format has been designed with pedagogical principles in mind. Visual aids and video are used in addition to text, as well as a 'pre-task' and reflection upon each topic. The activities at the end of each section enable learners to develop practical experience of each step of the actual archiving process, from planning to submission. There are also suggestions for further reading and resources to allow participants to engage with topics in more detail outside the scope of the course.

This resource is helpful for those undertaking their first language documentation project, particularly those coming from multidisciplinary teams who may not have much language documentation background to draw on, or for language community members who are carrying out documentation under their own steam. The discussion of progressive archiving is also helpful since it makes the whole enterprise of language documentation less intimidating by encouraging smaller, neater submissions to archives. This has the effect of encouraging language documentation since it reframes it as something more achievable and manageable than the commonly held perception of language documentation projects as huge, complex undertakings. In addition, considering Dwyer's urging to "do some good" – not only in the sense of compensating language communities for their contributions but also by adding value through education or empowerment – the certificate is a nice touch, particularly for community members collaborating in language documentation projects (Dwyer 2006). This, of course, depends on their ability to read English, which is perhaps the only drawback of this resource. However, since it is so new, having only been released in 2020, it cannot be too harshly criticised for not yet being available in multiple languages. Additionally, The Archive of the Indigenous Languages of Latin America (AILLA 2020) have already begun to make the video content for AFTF available in Spanish via their YouTube channel, and it is likely that this multilingual expansion is something that will continue to be developed in future. Overall, *Archiving for the future: Simple steps for archiving language documentation collections* is a well-designed and well-structured tool that easily achieves its stated goals of bringing together the main concepts and tools of contemporary language documentation into a single, accessible resource.

References

AILLA. 2020. *Tipos de arquivo*. AILLA Archive of the Indigenous Languages of Latin America. https://youtu.be/trl5EEVg950

Bird, Steven & Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language* 79(3). 557–582.

Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, et al. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal* 19. 43.

Dwyer, Arienne. 2006. Ethics and practicalities of cooperative fieldwork and analysis. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 31–66. Berlin, New York: Mouton de Gruyter.

Henke, Ryan & Andrea L. Berez-Kroeker. 2016. A Brief History of Archiving in Language Documentation, with an Annotated Bibliography. *Language Documentation & Conservation* 10. 411–457.

Holton, Gary. 2012. Language archives: They're not just for linguists any more. (Ed.) Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts & Paul Trilsbeek. *Language Documentation & Conservation* SP3: Potentials of language documentation: Methods, analyses, and utilization. 105–110.

Jukes, Anthony. 2011. Culture documentation and linguistic elicitation. In Nicholas Thieberger, Linda Barwick, Rosey Billington & Jill Vaughan (eds.), *Sustainable data from digital research: Humanities perspectives on digital scholarship*, 47–62. Melbourne: Custom Book Centre, University of Melbourne.

Kung, Susan Smythe, Ryan Sullivant, Elena Pojman & Alicia Niwabaga. 2020. *Archiving for the Future: Simple steps for archiving language documentation collections*. https://archivingforthefuture.teachable.com/p/home (22 July 2021).

Nathan, David. 2013. Progressive archiving: Theoretical and practical implications for documentary linguistics. Presented at the 3rd International Conference on Language Documentation and Conservation (ICLDC).

Nathan, David. 2014. Access and accessibility at ELAR, an archive for endangered languages documentation. (Ed.) David Nathan & Peter K. Austin. *Language Documentation and Description* 12. 187–208.

Nathan, David & Peter K. Austin. 2004. Reconceiving metadata: language documentation through thick and thin. *Language Documentation and Description* 2. 179–187.

Sullivant, Ryan. 2020. Archival description for language documentation collections. *Language Documentation & Conservation* 14. 520–578.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1). 160018.